

Quick and dirty – hvordan søges der i DanBib?

Informationsprofessionelles søgeadfærd i DanBib – kunne den være mere effektivt?

Af Kirsten Larsen

Er der en tilbøjelighed til at overføre søgestrategier fra søgemaskiner ukritisk til andre web-redskaber, selvom strategien ikke nødvendigvis er den bedste her?

Af Kirsten Larsen
Afdelingsleder, DBC
kl@dbc.dk

Eller endnu værre: Påvirker Google på den negative måde så alt, hvad der ikke kan findes på to sekunder nok ikke er værd at lede efter?

På DBC laver vi ind imellem analyser af log-filerne i bibliotek.dk eller DanBib – som regel med ret konkrete formål á la:

- bliver de løsninger vi laver brugt – fx Søg om forfatteren i bibliotek.dk?
- er der noget vi skal tage højde for i forbindelse med en ny søgemaskine?
- kan vi skære nogle søgekoder væk, så opdateringen bliver hurtigere?
- hvor mange 0-hit-søgninger er der? – hvad kan vi gøre for at mindske antallet (der hvor der burde have været fundet poster)?
- osv.osv.

I vinter lavede jeg forskellige undersøgelser i bibliotek.dk som forberedelse til arbejdet med den ny brugergrænseflade, som sættes i drift ca. 1. oktober i år. Samtidig kiggede jeg en smule på DanBib, fordi det jo altid er rart at sammenligne sådanne tal med noget andet. Desværre er der jo ikke mange der laver den slags undersøgelser i bibliografiske baser – næsten al forskning handler om søgning på nettet eller i fuldttekstbaser.

Bibliotek.dk

Her har vi en meget stor database med bibliografiske data af meget forskellige karakter og ringe muligheder for en for hver enkelt bruger meningsfuld relevansrangering. Der er store muligheder for støj i forhold til behov i fritekstsøgninger.

Derfor forsøger brugergrænsefladen at få oplysninger ud af brugerne, selvom de måske ikke umiddelbart tænker over, at de har dem. Og vi ser, at det faktisk lykkes – brugerne søger i høj grad i de bokse, som udfører en smallere søgning end fritekstsøgning – ligesom de tiltag vi laver i form af almindelige links til videresøgning, *Noget der ligner*, *Litteratur om forfatteren* i høj grad bruges.

Medvirkende til mønstret er naturligvis også, at mange bruger bibliotek.dk til ren verifikativ søgning. Det mere græsningsprægede behov tilgodeser en del brugere via deres lokale katalog – eller fysisk græsning på et bibliotek.

DanBib i netpunktsversion

Denne version af DanBib var den første web-version af DanBib, som skulle kunne klare sig uden en alternativ tegnbaseret version. (Der var tidligere webversioner, som blev brugt langt mindre – blandt andet fordi der var et tegnbaseret alternativ og vel også fordi funktionaliteten ikke var så omfattende). Det var derfor vigtigt under udviklingen af netpunkt.dk, at alle de vaner, som brugerne havde fra den tegnbaserede, nogenlunde smertefrit kunne videreføres til web-udgaven. I hvert fald så det samlet gav en bedre funktionalitet. Visning i web (bl.a. link til videresøgning og opslag i lokalkataloger) er jo nærmest ”automatisk” langt bedre i web-miljøet.

Den arbejdsgruppe af biblioteksfolk, som var med i udformningen, lagde vægt på, at man skulle have en boks, hvor der kunne skrives ren ccl, så søgekoder ikke var noget, der skulle hentes fra menuer eller lignende. Det er da også denne brugergrænseflade der bruges af 97 % af brugerne, selvom der er en menubase-

Tallene fra bibliotek.dk og DanBib ser pudsige ud sat over for hinanden:

Analyse af søgninger i bibliotek.dk (20.506 søgninger – 20. december 2004)

Fritekst	7%
Forfattersgning (boks og kammando)	34%
Videresøgning (link) på forfatter	5%
Titelsøgning	20%
Emnesøgning (boks og kammando)	11%
Videresøgning (link g ”Litteratur om.” på emne	15%
Resten (hver enkelt max 2%)	i alt 8 %

Analyse af 1 dags søgninger i DanBib via netpunkt (37.605 søgninger – 5.1. 2005)

Fritekst	48%
ti og lti	21%
fo og lfo	8%
is og nr	5%
em og lem	3%
dk	15%
En lang række koder: under 1%	i alt 8 %
En endnu længere række: ingen	

På mange måder kan man sige, at brugerne retter sig efter brugergrænsefladen

Nogle søgestrategier fra søgemaskiner er egnet til søgning i fuldtekstdokumenter og som oftest i ikke-formaterede dokumenter. De samme strategier er ikke nødvendigvis de hurtigste eller bedste i bibliografiske baser med formaterede metadata.

ret version også. Man kan også ved klik hente søgekoder fra hjælpebokse på søgefladen, men det er ikke mit indtryk, at det bruges meget. Det er den slags, som vi ikke kan se i loggen, så det ved brugerne mere om end jeg.

Og skriver de biblioteksansatte så ccl? – ja det kan unægtelig diskuteres – halvdelen af søgningerne er fritekstsøgninger uden søgekoder.

Og i forhold til de muligheder der er for trunkering (højre og venstre), maskering, operatører og 139 søgekoder, må man sige, at der er langt mellem snapsene = de mere avancerede søgninger.

Nu er der jo heller ikke nogen grund til at søge avanceret, hvis man kan finde det, man leder efter uden brug af andet end et ord eller to med et "og" imellem dem!

Hvad ledes der efter?

Der er ingen tvivl om, at en stor del af det, der søges efter i DanBib, er konkrete titler. Og der laves fornuftige søgninger – typisk kombinatoriske søgninger med et "og":

ivarsson og subtitling

lambron og menteurs

thomsen og ambulancerne

thuren og begyndere

Alle søgninger, som hurtigt bringer én frem til, hvad man leder efter, uden at det tager mere end et par sekunder at skrive søgningen. Her er der ingen som helst grund til at ændre søgestrategi – det er hurtigt og effektivt.

Imidlertid synes jeg, at studiet af logningen giver det indtryk, at der mellem alle disse relevante og hurtige verifikative søgninger også er en del, hvor der virkelig kæmpes for at finde et eller andet – og hvor det umiddelbart kunne se ud som om lidt mere brug af søgekoder eller kombinatorik eller begge dele, kunne gøre gavn. Andre søgninger finder givet det, der søges efter, men en anden strategi ville måske hurtigere bringe én frem til det ønskede.

Nogle søgestrategier fra søgemaskiner er egnet til søgning i fuldtekstdokumenter og som oftest i ikke-formaterede dokumenter. De samme strategier er ikke nødvendigvis de hurtigste eller bedste i bibliografiske baser med formaterede metadata.

Og i alle søgesystemer (søgemaskiner så vel som databaser) er der ting, der er tunge, ting der er "forbudt" eller stopord – eller ting man ikke kan. For professionelle brugere er det naturligvis en god ting at kende til disse ting i ens daglige arbejdsredskaber. I de redskaber vi bruger en gang om måneden eller mindre, foretrækker de fleste nok brugergrænseflader a la bibliotek.dk eller ganske mange fagdatabaser, hvor man ledes til den mest hensigtsmæssige søgemåde uden at skulle tænke sig om.

Det er jo altid farligt at give sig til at konstruere sig til hvilke søgebehov og formuleringer af disse, der ligger bag noget, man ser i en logfil. Forsøgene her skal kun ses som ideer til, hvordan man i mange situationer hurtigere kunne skaffe sig det, der ønskes.

I virkeligheden er jeg en del gange under afprøvningen af søgningerne blevet overrasket over, at svartiden eller resultatet ikke var værre – sommertiden kan dog betyde, at svartiderne er en del bedre end i januar, hvor søgningerne oprindeligt blev udført.

kvale og intervi?

kvale og interview og ma=bå

kvale og interwi?

kvale og kvalitative

Første søgning giver 26 poster – de bliver nok vist i et større format, for efter en del bladringer bliver man utålmodig og forsøger en ny strategi. En søgning på fo=kvale og em=intervi? og ma=bå giver en halvering af søgeresultatet og det er (formentlig) let at finde den rette post.

Søgekoderne er altså ikke nødvendige for at finde den rette post, man finder den blot hurtigere, og behøver derfor ikke at begive sig ud i alternative strategier undervejs.

Hvad leder vi efter her?

orkerede julestjerner

orkerede kors

orkerede kors

orkerede sommerfugle

orkerede?

orkerede?

orkis og julestjerner?

orkis og julestjerner?

orkis og kors?

orkis og kors?

orkis og sommerfugl?

orkis og sommerfugl?

Det ser ud som om, vi skal have fat på én, to eller tre bøger om orkis, der handler om henholdsvis julestjerner, kors og sommerfugle.

En søgning på em=orkis giver alle tre titler i ét hug:



Slutbrugere kan af og til ønske ”absolut alt”, men en professionel bruger ville nok ofte have glæde af at opdele i to strategier: én til at søge ”litteratur af ” og én til ”litteratur om”

En søgning: *jackie og chan* giver 70 poster (der er bladret en del, kan man se i logfilen...) – mon ikke en søgning fx *fo=chan jackie* eller *em=chan jackie*, som giver 20 poster, giver det, der ledes efter? Slutbrugere kan af og til ønske ”absolut alt”, men en professionel bruger ville nok ofte have glæde af at opdele i to strategier: én til at søge ”litteratur af ” og én til ”litteratur om”. (Jo, der er nogle videoer med Jackie Chan, man ikke finder med en *em*-søgning).

Ordregistre – langordsregistre

I DanBib har vi valgt at gøre det muligt at lave ordsøgning i alle felter – dvs. alle elementer i dataene lægges ud enkeltvis, så man frit kan kombinere, selvom man kun kender et fornavn og et eller andet ord i en titel eller emnet. Eller man kender to ord, som står ved siden af hinanden, men man ved ikke hvor.

Ligeledes ligger en lang række data også i langordsregistre (phrase indexes), som muliggør dels meget hurtig søgning, hvis man har præcise data, dels skanning i registre, så man ofte kan bladre sig frem til det ønskede.

Men det betyder også, at visse typer søgninger er særdeles tidkrævende – fra logningen:

den gode leder er et dydsmønster

en rejse til det gode lederskab

det athenske demokrati i samtidens og eftertidens syn

the organic chemistry of drug synthesis

the declaration of independence and the constitution of the united states

Alle disse søgninger er foretaget i fritext.

De lader sig udføre i Danbib – i hvert fald på en rolig sommerdag, hvor belastningen ikke er så stor. Der er intet ”galt” i at søge sådan. Men fx den sidste kan udføres på 3 sekunder med søgekoden *lti* – det tog (stadig på en stille sommerdag) 24 sekunder, som den er skrevet.

Har man præcise titler, kan man lige så godt udnytte fordelene ved langordsregistre. Kan man ikke huske koderne, står de i den gule boks på forsiden og et enkelt klik på dem, kopierer koden op i indtastningsfeltet.

Google-syndromet?

Endnu et eksempel fra logfilen: *den store slæderejse gyldendalske nyt nordisk*

Er der en tilbøjelighed til at overføre søgestrategier fra søgemaskiner ukritisk til andre web-redskaber, selvom strategien ikke nødvendigvis er den bedste her? Eller endnu værre: Påvirker Google på den negative måde, så alt, hvad der ikke kan findes på to sekunder, nok ikke er værd at lede efter?

Anbefalinger

Den lille stikprøve giver anledning til at repetere disse anbefalinger:

1. Pluk et ”godt” ord i titlen og kombiner med en lige så god del af navnet – så er fritext lynhurtig og effektiv.

2. Hvis man har kendte titler med mange småord, er søgning med *lti* (slut gerne med et spørgsmålstegn for en sikkerheds skyld – undertitler kan drille) eller en kombinatorisk søgning med ”gode” ord langt hurtigere end en lang frasesøgning.

3. Skal man lave emnesøgninger, kan det virkelig betale sig at lære nogle søgekoder eller slå dem op, når behovet er der.

I forbindelse med udarbejdelse af emnehierarkiet til bibliotek.dk, konkluderede de indekserer, der har lavet hierarkiet, at 5 søgekoder kunne klare det meste: **ok, dk, em, ke, df**.

Det er to klassifikationssøgekoder: **dk** til alle dk5-klassemærker, og **ok** til dk-opstillingsklassemærke – særdeles nyttige, når man skal finde nogle centrale titler uden at bruge for megen tid.

Og tre emnesøgekoder: den brede **em**, som alle vel kan lære, **ke** som er kontrollerede emnesøgekoder og **df** som er den smalle kode til faglitteratur indekseret med DBC-termer – god når der er meget litteratur.

Alle koder kan ses i hjælpeboksen på DanBib's forside eller i det hjælpevindue med flere emnesøgekoder, som er ét klik fra det.

Når man kigger emnehierarkiet igennem, er det tydeligt, at den hyppigst anvendte søgestrategi til at finde central dansk litteratur om et emne, er en kombination af et klassemærke (med **dk** eller **ok**) og et emneord (**em**, **ke** eller **df**). I logningen fra DanBib (mere end 37.000 søgninger) var de tre af koderne (**ok**, **ke** og **df**) slet ikke repræsenteret!

Hvad kan vi gøre for at gøre det lettere at lave effektive søgninger i DanBib?

Skal vi lave et emnehierarki i DanBib ligesom i bibliotek.dk? – hvor biblioteksmennesker med dybtgående viden om data og søgning har lavet søgninger på de mest almindelige emner – og til brug i den travle hverdag, når brugeren spørger om noget, der ikke lader sig finde i éns eget bibliotekskatalog?

Skal vi have ”dagens tip” på brugergrænsefladen? – eller i en danbib-tip-info?

Prioriteres den løbende vedligeholdelse af søgekompetencer for lavt i bibliotekerne?

Andre ideer?

Skriv gerne til **kl@dbc.dk**, hvis der er noget du mener, vi kan gøre fra DBCs side.